

SPECIFICATION AMENDMENTS

On page 7, line 16 to page 8, line 2, please delete the Description of Figures section in its entirety, as follows:

~~—DESCRIPTION OF THE FIGURES~~

~~Figure 1 shows a schematic representation of one embodiment of the systems of the present invention.~~

~~Figure 2 shows a schematic representation of a conference bridge configuration in one embodiment of the present invention.~~

~~Figure 3 shows a schematic representation of a processor configuration in one embodiment of the present invention.~~

~~Figure 4 shows a representation of a media player in one embodiment of the present invention.~~

~~Figure 5 shows a schematic representation of system connectivity in one embodiment of the present invention.~~

~~Figure 6 shows a schematic representation of a talk show format using the systems and methods of the present invention.~~

~~Figure 7 shows a schematic representation of a corporate meeting using the systems and methods of the present invention.~~

~~Figure 8 shows a schematic representation of the generation of translation and sub-titles for video using the systems and methods of the present invention.~~

On page 7, line 17 to page 8, line 2, please insert the following replacement paragraph:

—The present invention provides systems for processing media events to generate text from an audio component of a media event and to process, as desired, and deliver the text to a viewer. ~~One preferred embodiment of the systems of the present invention is diagrammed in Figure 1. Figure 1 shows a number of components, including optional components, of the systems of the present invention. In this embodiment~~ In some embodiments, the audio information of a media event is transferred to a conference bridge. Audio information received by the conference bridge is then sent to one or more other components of the system. For example, audio information may be sent to a speech-to-text converter (e.g., a

captionist/transcriptionist and/or voice recognition software) where the audio is converted to text. The media information received by the conference bridge may also be sent directly to a processor that encodes the audio for delivery to a viewer (*e.g.*, compresses the audio and/or video components of multimedia information into streaming data for delivery to a viewer over a public or private electronic communication network). Text information that is generated by the speech-to-text converter is also sent to the processor for delivery to a viewer. In preferred embodiments, the text information is encoded in a separate delivery stream than the audio or video components of the multimedia information that is sent to a viewer. The text information, as desired, can be translated into one or more different languages. For example, ~~in Figure 1~~ in some embodiments, the encoded text stream is translated using a real-time language translator (*e.g.*, SysTran, Enterprise).—

On page 19, line 31 to page 20, line 18, please insert the following replacement paragraph:

~~—An example of a conference bridge that finds use in an interactive talk-show format is diagrammed in Figure 2.~~ In some embodiments of the present invention, the conference bridge is used in an interactive talk-show format. ~~In this~~ For example, multimedia information generated at a live event is transmitted to the conference bridge. The multimedia information includes audio from a moderator and participants of the live event. Audio information can also be received from one or more remote recipients. Viewers (*e.g.*, call-in viewers) of the talk-show can also send audio information to the conference bridge. As desired, the information content from the call-in viewers can be screened to determine if it is appropriate to disseminate to other viewers or participants. In such embodiments, a call-in screener is connected to the conference bridge such that the call-in screener monitors the call-in audio from the viewers prior to it being heard or viewed by other viewers or participants. The conference bridge can be configured to allow different levels of access and information processing. For example, the event participant audio information can automatically be processed to text, while the call-in viewer audio is originally directed to a private call-in virtual conference, monitored, and only sent to the live virtual conference for text conversion if approved by the screener. Information that is to be converted to text is sent to a speech-to-text converter. The speech-to-text converter need not receive the video of the live event, but can simply be sent the audio (*e.g.*, through the conference

bridge) that is to be converted to text. Additional participants may also be connected to the conference bridge including a system administrator or operator. The control of the conference bridge can be operated directly or over a communications network. For example, all of the moderator, participant, and administrator functions can be controlled over the World Wide Web.—

On page 23, lines 23-30, please insert the following replacement paragraph:

~~—As shown in Figure 3~~ In some embodiments, multimedia information is received by a processor through a conference bridge and/or from a speech-to-text converter and converted to an appropriate format to allow useful delivery to one or more viewers. For example, in some embodiments of the present invention, streaming media is used to provide audio, video, and text to viewers. In such embodiments, the processor encodes one or more information streams from the audio and/or video information of the multimedia information. The processor also encodes (e.g., separately) a text stream. The text and multimedia information are then sent, directly or indirectly, to one or more viewers.—

On page 29, lines 9-26, please insert the following replacement paragraph:

~~—An example of a media player that finds use with the present invention is shown in Figure 4. This~~ In some embodiments, the media player contains a viewer screen for viewing video and a separate text box. ~~Figure 4 shows the use of the media player in conjunction with the motion picture “Sleepless in Seattle.”~~ The video and audio are controlled by ~~the~~ a panel under the video screen that allows for starting, stopping, fast forward, reverse, and volume control. The text box displays the name of the speakers, or their title, and provides a text transcript of their spoken audio. Controls under the text box allow the text to be viewed in different languages and allow the audio to be changed to the language selected. The viewer using the media player can select the option “view transcript” which opens a separate text box containing the current accumulative transcript in the language selected. This text box can be configured to allow text to be edited, copied, printed, searched and otherwise manipulated. The top of the media player also includes a box for the viewer to enter comments/questions and send them back to a question queue on the database. The present invention provides a web-based control for event screening, approval and prioritizing of viewer entered comments/questions. In

this case, comments/questions are entered as text and are processed through the systems of invention, although they could also be sent as voice-over-IP audio, public switched network (telephone) audio, email, or in any other desired format. The systems of the present invention are also configured to allow other viewers to view event approved comments/questions.—

On page 30, lines 3-23, please insert the following replacement paragraph:

~~—Figure 5 shows one example of a system configuration of the present invention. Audio~~
In some embodiments of the present invention, audio information is passed from a conference bridge to a speech-to-text converter. The multimedia information from the conference bridge and the text information from the speech-to-text converter are sent to a processor where the media and text are separately encoded into streaming information. The processor is connected to a web server (*e.g.*, a web server comprising FTP, IIS, and C52K servers), databases, and streaming media servers through a network (*e.g.*, a local area network (LAN)). Streaming audio and video information are sent from the processor to the streaming media server and streaming text is sent to a Java applet running on the viewers' browser. A media player (*e.g.*, custom SPECHE BOX software with embedded media player, SPECHE COMMUNICATIONS) viewable by a viewer receives the text and multimedia information and displays the multimedia performance and text to a viewer. The viewer can opt to "view transcript," which sends a request to an FTP server to supply the full transcript (*e.g.*, the full transcript as generated as of the time the viewer selected the option) to the viewer. The viewer can also send information (*e.g.*, comments/questions) back to the processor. ~~In the embodiment shown in Figure 5~~ some embodiments, a data control system (*e.g.*, one or more computers comprising a processor and/or databases) allows the viewer to register, provides schedule information on the event, and receive viewer question information. Storage of viewer information in a database at registration allows viewer preferences to be determined and stored so that delivered content is correct for each individual. Customer registration and event scheduling information is also stored in the database to automate and control event operations using the ~~Rob-Cop~~ Robo-Cop (Expert System Systems), and to administrate the transaction / business relationship.—

On page 32, lines 3-17, please insert the following replacement paragraph:

—A similar process can be applied to provide translated text (e.g., subtitles) for television programming or any other multimedia presentation where it may be ~~desireable~~ desirable to have language translations applied (e.g., video presentations on airlines). ~~One embodiment for video translation and sub-titling is shown in Figure 8. In this figure~~ In some embodiments, an original video with audio in a first language (e.g., English) is processed into encoded audio and video (e.g., in .WMA and .WMV file formats). In some embodiments, encoded audio and low quality encoded video are sent (e.g., via Web FTP) to a conference bridge of the present invention, where audio is converted to text by a speech-to-text converter and translated by a language translator using methods described above. The translated text (e.g., in the form of a translated script) is then sent to a foreign territory where the translated information is used to re-dub the video with foreign language voice over. Text information (in one or more different languages) may also be sent to a video studio to prepare sub-titles in any desired language (e.g., as a final product or for preparing an intermediate video to be sent to the foreign territory to prepare a re-dubbed video). The physical location of any of the systems does not matter, as information can be sent from one component of the system to another over communication networks.—

On page 32, line 20 to page 33, line 12, please insert the following replacement paragraph:

—Many newsworthy events (*e.g.*, political speeches, etc.), business proceedings (*e.g.*, board meetings), and legal proceedings (*e.g.*, trials, depositions, etc.) benefit from or require the generation of text transcripts (and optional translations) of spoken language. The systems and methods of the present invention provide means to generate real-time (or subsequent) text transcripts of these events. The text transcripts can be provided so as to allow full manipulation of the text (*e.g.*, searching, copying, printing, etc.). For example, news media personnel can receive real-time (or subsequent) transcripts of newsworthy speeches, allowing them to select desired portions for use in generating their news reports. A major advantage of using the systems and methods of the present invention is that the user of the text information need not be present at the location where the event is occurring. Virtual business meetings and legal proceedings are possible, where each of the participants receives a real-time (or subsequent) copy of the text of the proceeding, as it occurs. Non-live event transcripts/translations are

created after the audio from a prior live event has been recorded for subsequent playback for transcription and translation by captionist/transcriptionist. ~~One embodiment of such an application is illustrated in Figure 7.~~ A In some embodiments of the present invention, a potential corporate customer registers (and is approved) on a web site and pre-buys a block of minutes (or hours) of transcription (and optionally translation) services. During a corporate meeting (*e.g.*, Board Meeting), the meeting chairperson (*e.g.*, on a quality speakerphone) calls into the systems of the present invention and enters their service access code for the transcription / translation services pre-purchased. The meeting participants conduct a normal meeting, speaking their name prior to participation. At the end of the meeting, the chairperson simply hangs-up the phone. Within a required duration (predetermined as a service option), the transcripts (in selected languages) are e-mailed or otherwise delivered to the designated address (or made available on a secured web sight). The customer's account is decremented, and they are notified when service time reaches a pre-determined balance. This service would also make the recorded audio available in the original (and optionally translated) languages.—

On page 34, line 5 to page 35, line 3, please insert the following replacement paragraph:

—The systems and methods of the present invention provide for interactive events involving viewers located in different areas. These interactive events include talk-show formats, debates, meetings, and distance learning events. In some embodiments, interactive events are conducted over the Internet. ~~An example of a talk show format is provided in Figure 6.~~ An For example, an event moderator can control the system through a web-based interface so that participants need not be burdened with equipment shipping, training, and maintenance. Participants can be anywhere in the world allowing for virtual web debates, distance instruction and education in which interaction is critical to the learning process, and intra organizational communication within large organizations with multiple offices in various foreign countries. Any event that can benefit from question and answer interactivity with an offsite audience finds use with the systems and methods of the present invention. Participant questions can be directed over the telephone or typed as in a chat format and can be viewed by all other participants in real time and/or after the fact. The systems and methods of the present invention provide dramatic flexibility for involving participants who speak different languages. The systems and methods of the present invention translate all viewer comments and questions from their selected language to

that of the screener (or moderator) to facilitate screening and prioritizing. All comments and questions entered (and approved by the screener) in various languages by all viewers are translated to the selected language of each viewer. This approach insures that all viewers gain the greatest benefit from an event, by interacting in their selected language for: streaming transcript, accumulative complete transcripts, audio dialogue, and comments / questions entered and received. In ~~the embodiment shown in Figure 6~~ some embodiments, the web presenter accesses a database of the present invention to register and schedule the event. The database can also be used to store an image file of the presenter, presentation files (e.g., POWERPOINT presentation files), and a roster of information pertaining to invited participants. The information in the database may be updated during the presentation. For example, questions from viewer participants and responses may be stored on the database to allow them to be viewed at the request of any of the participants. Questions from viewer participants may be received aurally using voice-over IP technology. These questions are directed to the conference bridge, with the audio being converted to text by a speech-to-text converter and the text information and/or corresponding audio information being routed to a processor for encoding as text and/or multimedia information streams, as well as storage in the database. At the request of any participant, the questions may be viewed as text and/or audio in any desired language.—